ASYMPTOTIC EXPONENTIALITY OF THE DISTRIBUTION OF FIRST EXIT TIMES FOR A CLASS OF MARKOV PROCESSES WITH APPLICATIONS TO QUICKEST CHANGE DETECTION

MOSHE POLLAK
The Hebrew University of Jerusalem
Department of Statistics
Mount Scopus
Jerusalem 91905, Israel
msmp@mscc.huji.ac.il

ALEXANDER G. TARTAKOVSKY
University of Southern California
Department of Mathematics
3620 S. Vermont Ave, KAP-108
Los Angeles, CA 90089-2532, USA
tartakov@usc.edu

Submitted to Probability Theory and Its Applications, March 2007

Abstract

We consider the first exit time of a nonnegative Harris-recurrent Markov process from the interval $[0, A]$ as $A \to \infty$. We provide an alternative method of proof of asymptotic exponentiality of the first exit time (suitably standardized) that does not rely on embedding in a regeneration process. We show that under certain conditions the moment generating function of a suitably standardized version of the first exit time converges to that of $\mathrm{Exponential}(1)$, and we connect between the standardizing constant and the quasi-stationary distribution (assuming it exists). The results are applied to the evaluation of a distribution of run length to false alarm in change-point detection problems.

Keywords and Phrases: Markov Process, Stationary Distribution, Quasi-stationary Distribution, First Exit Time, Asymptotic Exponentiality, Change-point Problems, CUSUM Procedures, Shiryaev-Roberts Procedures.

1. Introduction

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and $\{X(n)\}$, $n = 0, 1, 2, \dots$ be a discrete-time nonnegative Harris-recurrent Markov process defined on this space. The limiting distribution as $A \to \infty$ of the suitably standardized first exit time of the process from the interval $[0, A]$ turns out often to be exponential.

The standard method for proving this asymptotic exponentiality is to try to find a version of the process that is regenerative (cf. Glasserman and Kou, 1995 and Asmussen, 2003). The heuristic behind this is that since the process is Harris-recurrent, it returns to a given set over and over again, and thus creates "cycles" that are "almost independent." Hence, the first cycle in which $X(n)$ exceeds A is approximately geometrically distributed, and if the expected length of a cycle

is finite and the probability of exceeding A in a given cycle tends to 0 as $A \to \infty$, then, suitably standardized, the asymptotic distribution of the first exit time is exponential.

In this paper, we make a connection between the standardization constant and the quasi-stationary distribution. Our method of proof is a coupling argument. Although less general as a method for proving asymptotic exponentiality than the regeneration argument, we believe that our method is of interest in its own right. This notwithstanding, the regeneration argument seems to be widely unknown in the statistics community, and ought to be publicized.

The paper is organized as follows. In Section 2, we present the main result that states that the limiting distribution of the suitably standardized version of the first exit time as $A \to \infty$ is $\mathrm{Exponential}(1)$ and that the moment generating function converges to that of $\mathrm{Exponential}(1)$, which implies that the convergence is in L^p for all $p \ge 1$. The proof is given in Section 3. We make a few remarks in Section 4. In Section 5, we give examples and describe applications to the evaluation of the distribution of the run length to false alarm for several change detection procedures.

2. Main Results

Let $\{X(n)\}_{n=0}^{\infty}$ be a discrete-time Harris-recurrent Markov process with state space $[0,\infty)$ and stationary transition probabilities. Let $\mathbf{P}^x$ denote the probability measure for the process when it starts at x (i.e., $X(0) = x$), and let $\mathbf{P}^G$ denote the probability measure when the initial state is distributed according to the distribution G.

Definition. We call the process stochastically monotone if $\mathbf{P}^x(X(1) \ge y)$ is non-decreasing and right-continuous in x for all y.

We will be interested in the behavior of the first exit time of $X(n)$ from the interval $[0,A]$ when $X(n)$ starts at $x \in [0,A)$, i.e., of the stopping time

$$N_A^x = \min\{n \ge 1 : X(n) > A\}, \quad X(0) = x, \tag{2.1}$$

where $0 \le x < A$ and A is a positive finite threshold, assuming that the Markov process $X(n)$ is stochastically monotone and Harris-recurrent.

The following theorem is the main result of the paper.

Theorem 1. *Let $X(n)$, $n = 0,1,2,\dots$ be a stochastically monotone Harris-recurrent Markov process with state space $[0,\infty)$ and stationary transition probabilities such that:*

C1. *The stationary distribution $H(y) = \lim_{n\to\infty} \mathbf{P}^x\{X(n) \le y\}$ exists and its support is $[0,\infty)$.*

C2. *The quasi-stationary distribution $H_A(y) = \lim_{n\to\infty} \mathbf{P}^x\{X(n) \le y | N_A^x > n\}$ exists for all $0 \le x < A$ and for all $0 < A < \infty$.*

Let $p_A = \mathbf{P}^{H_A}\{X(1) > A\}$.

Then:

(i) *The distribution of $p_A N_A^x$ is asymptotically $\mathrm{Exponential}(1)$ as $A \to \infty$ for all fixed $x \in [0,\infty)$.*

(ii) *The moment generating function $\mathbf{E}\exp\{t p_A N_A^x\}$ of $p_A N_A^x$ converges to $1/(1-t)$ as $A \to \infty$ for all fixed $x \in [0,\infty)$. In particular, it follows that*

$$\lim_{A\to\infty} p_A \mathbf{E} N_A^x = 1 \quad \textit{and} \quad \lim_{A\to\infty} \mathrm{Variance}\{p_A N_A^x\} = 1.$$

Conditions C1 and C2 hold in a variety of scenarios. See corresponding remarks in Section 4 and examples in Section 5.

We begin with a heuristic argument. A formal proof requires several auxiliary results and is given in Section 3.

Write $N_A^{H_A}$ for the stopping time when the process $X(n)$ starts at a random point $X(0) = \xi$ in $[0, A]$ that has a quasi-stationary distribution H_A, i.e., $\mathbf{P}(\xi \le y) = H_A(y)$. Then $\mathbf{P}^{H_A}(X(n) > A | N_A^{H_A} \ge n) = p_A$ for all $n \ge 1$, and, therefore, the distribution of $N_A^{H_A}$ is geometric with the parameter p_A for all $A > 0$. Further, under conditions C1 and C2, the probability p_A goes to 0 as $A \to \infty$, which implies that $p_A N_A^{H_A}$ converges weakly to $\mathrm{Exponential}(1)$ as $A \to \infty$. Intuitively, the asymptotic behavior of the stopping time N_A^x for every fixed point x is similar to that of $N_A^{H_A}$. Mathematical details are presented in the next section.

3. Proof

In order to prove Theorem 1 we need the following lemmas. We use the notation of the previous section, and we assume that the conditions of Theorem 1 are satisfied.

Lemma 1. *The quasi-stationary distribution*

$$H_A(y) = \lim_{n \to \infty} \mathbf{P}^x \left\{ X(n) < y | N_A^x > n \right\}$$

converges to the stationary distribution $H(y)$ at all continuity points y of H.

Proof. Follows directly from Theorem 1 of Pollak and Siegmund (1986). □

Recall that $N_A^{H_A}$ is the stopping time (2.1) when the Markov process $X(n)$ starts from the random point that has the quasi-stationary distribution H_A, i.e., $X(0) \sim H_A$.

Lemma 2. *The distribution of $N_A^{H_A}$ is $\mathrm{Geometric}(p_A)$, where $p_A = \mathbf{P}^{H_A}\{X(1) > A\}$. Hence $p_A \mathbf{E} N_A^{H_A} = 1$ and $p_A N_A^{H_A}$ converges in distribution to $\mathrm{Exponential}(1)$ as $A \to \infty$.*

Proof. Since the Markov process is Harris-recurrent, there is no absorbing state, so that $\mathbf{P}(N_A^{H_A} = \infty) = 0$. Therefore, the geometric property of $N_A^{H_A}$ is obvious. Lemma 1 and the assumption that the support of H is $[0, \infty)$ guarantee that $p_A \xrightarrow[A \to \infty]{} 0$. □

Lemma 3. *Let $X^x(n)$ denote a process that starts from x and has the same transition probabilities as $X(n)$. Let $0 \le x < y < \infty$. There exists a sample space with $X^x(n)$ and $X^y(n)$ such that $X^y(n) \ge X^x(n)$ for all $n \ge 1$.*

Proof. Clearly $X^y(1)$ is stochastically larger than $X^x(1)$, so that one can construct a sample space where $X^y(1) \ge X^x(1)$. To complete the proof, continue by induction on n. □

Lemma 4. *Let $0 \le x < y < \infty$. Let $\tilde{X}^x(n)$ and $\tilde{X}^y(n)$ be independent Markov processes started at x and y respectively, both having the same transition probabilities as $X(n)$. Then*

$$\mathbf{P}\left\{ \tilde{X}^x(n) > \tilde{X}^y(n) \text{ for at least one value of } n \right\} = 1. \tag{3.1}$$

Proof. Let $0 < \varepsilon < 1/4$ and $y \le B < \infty$ be such that $H\{(B,\infty)\} = \varepsilon$. Let w_ε be such that

$$\left|\mathbf{P}\left\{\tilde{X}^B(w_\varepsilon) \le z\right\} - H(z)\right| < \varepsilon \quad \text{for all } z$$

and

$$\left|\mathbf{P}\left\{\tilde{X}^0(w_\varepsilon) \le z\right\} - H(z)\right| < \varepsilon \quad \text{for all } z.$$

By virtue of Lemma 3,

$$\left|\mathbf{P}\left\{\tilde{X}^x(w_\varepsilon) \le z\right\} - H(z)\right| < \varepsilon \quad \text{for all } z.$$

Write m for the median of the stationary distribution H. Obviously,

$$\mathbf{P}\left(\{B \ge \tilde{X}^x(w_\varepsilon) \vee \tilde{X}^y(w_\varepsilon)\} \setminus \{B \ge \tilde{X}^x(w_\varepsilon) \ge m, \tilde{X}^y(w_\varepsilon) \le m\}\right)$$
$$\le (1-\varepsilon)^2 - (\tfrac{1}{2}-\varepsilon)^2$$

and

$$(\tfrac{1}{2}-2\varepsilon)^2 < (\tfrac{1}{2}-2\varepsilon)(\tfrac{1}{2}-\varepsilon) \le \mathbf{P}\left\{B \ge \tilde{X}^x(w_\varepsilon) \ge m, \tilde{X}^y(w_\varepsilon) \le m\right\} \le (\tfrac{1}{2}+\varepsilon)^2.$$

Similarly, for any $j \ge 2$ when $u < v$

$$(\tfrac{1}{2}+\varepsilon)^2 \ge \mathbf{P}\left\{\tilde{X}^x(jw_\varepsilon) \ge m, \tilde{X}^y(jw_\varepsilon) \le m | \tilde{X}^x((j-1)w_\varepsilon) = u, \tilde{X}^y((j-1)w_\varepsilon) = v\right\}$$
$$\ge (\tfrac{1}{2}-2\varepsilon)^2$$

and

$$\mathbf{P}\left(\{B \ge \tilde{X}^x(jw_\varepsilon) \vee \tilde{X}^y(jw_\varepsilon)\} \setminus \{B \ge \tilde{X}^x(jw_\varepsilon) \ge m, \tilde{X}^y(jw_\varepsilon) \le m\}\right)$$
$$\le (1-\varepsilon)^2 - (\tfrac{1}{2}-\varepsilon)^2 = \frac{3}{4} - \varepsilon.$$

Let $T_B = \min\left\{j : \tilde{X}^x(jw_\varepsilon) \vee \tilde{X}^y(jw_\varepsilon) > B\right\}$.

Using previous inequalities, we obtain

$$\mathbf{P}\left\{B \ge \tilde{X}^x(jw_\varepsilon) \ge \tilde{X}^y(jw_\varepsilon) \text{ for some } 1 \le j < T_B\right\} \ge \left(\tfrac{1}{2}-2\varepsilon\right)^2 \sum_{i=0}^{\infty} \left(\tfrac{3}{4}-\varepsilon\right)^i$$
$$= \frac{\left(\frac{1}{2}-2\varepsilon\right)^2}{1-\left(\frac{3}{4}-\varepsilon\right)}$$
$$= \frac{\left(\frac{1}{2}-2\varepsilon\right)^2}{\frac{1}{4}+\varepsilon}.$$

Letting $\varepsilon \to 0$ completes the proof. $\square$

Lemma 5. *Using the same notation as in Lemma 4,*

$$\mathbf{P}\left(\tilde{X}^x(\ell) \geq \tilde{X}^y(\ell) \text{ for some } \ell \leq n\right) \xrightarrow[n\to\infty]{} 1$$

uniformly in $0 \leq x < y \leq B$.

Proof. This follows directly from Lemma 4 and its proof. □

Lemma 6. *Let* $\varepsilon > 0$ *and let* $0 < B < \infty$ *be such that* $H\{(B,\infty)\} < \varepsilon$. *Let* $B \leq A < \infty$. *Then* $H_A\{(B,A)\} < \varepsilon$.

Proof. The lemma follows from the fact that $H_A(y) \geq H(y)$ for all $y \geq 0$ (cf. Theorem 1 of Pollak and Siegmund, 1986). □

Proof of Theorem 1 (i). Let $N_A^{H_A} = \min\{n : X(n) > A\}$ where $X(0) \sim H_A$. By Lemma 2, $N_A^{H_A} \sim \mathrm{Geometric}(p_A)$ and

$$\lim_{A\to\infty} \mathbf{P}\left(p_A N_A^{H_A} > s\right) = e^{-s}, \quad s > 0.$$

Let $\varepsilon > 0$. Let $0 < B < \infty$ be such that $H\{(B,\infty)\} < \varepsilon$. Using the notation of Lemma 4, let $0 < q_B < \infty$ be such that

$$\mathbf{P}\left(\tilde{X}^0(n) \geq \tilde{X}^B(n) \text{ for some } n \leq q_B\right) > 1 - \varepsilon. \tag{3.2}$$

By virtue of Lemma 1 and Lemma 2, there exists A_ε such that for all $A \geq A_\varepsilon$

$$|H_A(x) - H(x)| \leq \varepsilon \quad \text{for all } 0 \leq x \leq B \tag{3.3}$$

and

$$\left|\mathbf{P}\left(p_A N_A^{H_A} > s\right) - e^{-s}\right| \leq \varepsilon \quad \text{for all } 0 \leq s < \infty. \tag{3.4}$$

Because the support of H is $[0,\infty)$, it follows from (3.3) that $p_A q_B \xrightarrow[A\to\infty]{} 0$.

Next, we construct the following sample space. Let $\hat{X}^0(n)$ be a Markov process (with transition probabilities as $X(n)$) starting at 0 and let $\hat{X}^B(n)$ be a Markov process starting at B such that they are independent until the first time that $\hat{X}^0(n) \geq \hat{X}^B(n)$. Denote this time by τ. After τ, let $\hat{X}^0$, $\hat{X}^B$ be such that $\hat{X}^0(n) \geq \hat{X}^B(n)$ for all $n \geq \tau$. (This construction is feasible by virtue of Lemma 3 and Lemma 4.)

By virtue of equation (3.2), $\mathbf{P}(\tau \leq q_B) > 1 - \varepsilon$. Denote

$$\hat{N}_A^0 = \min\left\{n \geq 1 : \hat{X}^0(n) > A\right\} \quad \text{and} \quad \hat{N}_A^B = \min\left\{n \geq 1 : \hat{X}^B(n) > A\right\}.$$

Note that N_A^x is stochastically larger than N_A^y if $x < y$.

Now, fix $0 \le s < \infty$ and let A_B be large enough so that $p_A q_B < s$ for all $A \ge A_B$. Then we have the following chain of equalities and inequalities:

$$
\begin{aligned}
\mathbf{P}\left(p_A N_A^B > s\right) &= \mathbf{P}\left(p_A \hat{N}_A^B > s\right) \\
&\ge \mathbf{P}\left(p_A \hat{N}_A^B > s, \tau \le q_B\right) \\
&= \mathbf{P}\left(p_A \hat{N}_A^B > s, p_A\tau \le p_A q_B < s\right) \\
&= \mathbf{P}\left(p_A \hat{N}_A^B > s, p_A\tau \le p_A q_B < s, \hat{N}_A^B > \tau\right) \\
&\ge \mathbf{P}\left(p_A \hat{N}_A^0 > s, p_A\tau \le p_A q_B < s, \hat{N}_A^0 > \tau\right) \\
&= \mathbf{P}\left(p_A \hat{N}_A^0 > s, p_A\tau \le p_A q_B < s\right) \\
&= \mathbf{P}\left(p_A \hat{N}_A^0 > s, \tau \le q_B\right) \\
&\ge \mathbf{P}\left(p_A \hat{N}_A^0 > s\right) - \mathbf{P}\left(\tau > q_B\right) \\
&\ge \mathbf{P}\left(p_A \hat{N}_A^0 > s\right) - \varepsilon \\
&= \mathbf{P}\left(p_A N_A^0 > s\right) - \varepsilon.
\end{aligned}
\tag{3.5}
$$

On the other hand,

$$
\begin{aligned}
\mathbf{P}\left(p_A N_A^B > s\right) &= \mathbf{P}\left(p_A N_A^{H_A} > s | X(0) = B\right) \\
&\le \mathbf{P}\left(p_A N_A^{H_A} > s | X(0) \le B\right) \\
&= \frac{\mathbf{P}\left(p_A N_A^{H_A} > s, X(0) \le B\right)}{\mathbf{P}\left(X(0) \le B\right)} \\
&= \frac{\mathbf{P}\left(p_A N_A^{H_A} > s, X(0) \le B\right)}{H_A([0,B])} \\
&\le \frac{\mathbf{P}\left(p_A N_A^{H_A} > s\right)}{H_A([0,B])}.
\end{aligned}
$$

Since by the definition of B and Lemma 6, $H_A([0,B]) \ge 1-\varepsilon$, and by equation (3.4), $P(p_A N_A^{H_A} > s) \le e^{-s} + \varepsilon$, we obtain

$$
\mathbf{P}\left(p_A N_A^B > s\right) \le \frac{e^{-s} + \varepsilon}{1 - \varepsilon}. \tag{3.6}
$$

Also, since $P(X(0) \ge 0) = H_A([0,A]) = 1$,

$$
\begin{aligned}
\mathbf{P}\left(p_A N_A^0 > s\right) &= \mathbf{P}\left(p_A N_A^{H_A} > s | X(0) = 0\right) \\
&\ge \mathbf{P}\left(p_A N_A^{H_A} > s | X(0) \ge 0\right) \\
&= \frac{\mathbf{P}\left(p_A N_A^{H_A} > s, X(0) \ge 0\right)}{P(X(0) \ge 0)} \\
&= \mathbf{P}\left(p_A N_A^{H_A} > s\right) \\
&\ge e^{-s} - \varepsilon,
\end{aligned}
\tag{3.7}
$$

where the last inequality follows from equation (3.4).

Putting (3.5) and (3.7) together yields

$$\mathbf{P}\left(p_A N_A^B > s\right) \geq e^{-s} - 2\varepsilon, \tag{3.8}$$

and putting (3.5) and (3.6) together obtains

$$\mathbf{P}\left(p_A N_A^0 > s\right) \leq \frac{e^{-s}+\varepsilon}{1-\varepsilon} + \varepsilon. \tag{3.9}$$

Since for all $0 \leq x \leq B$,

$$\mathbf{P}\left(p_A N_A^B > s\right) \leq \mathbf{P}\left(p_A N_A^x > s\right) \leq \mathbf{P}\left(p_A N_A^0 > s\right), \tag{3.10}$$

equations (3.8)–(3.10) imply that

$$e^{-s} - 2\varepsilon \leq \mathbf{P}\left(p_A N_A^x > s\right) \leq \frac{e^{-s}+\varepsilon}{1-\varepsilon} + \varepsilon \quad \text{for all } 0 \leq x \leq B.$$

Finally, fix x and let $\varepsilon \to 0$, so that ultimately $B > x$. This completes the proof of Theorem 1 (i).

PROOF OF THEOREM 1 (II). Since $N_A^{H_A}$ is distributed $\mathrm{Geometric}(p_A)$, $p_A N_A^{H_A}$ has a moment generating function

$$M_A^{H_A}(t) = \mathbf{E} e^{t p_A N_A^{H_A}}, \quad t < 1,$$

and it is easy to see that

$$M_A^{H_A}(t) \xrightarrow[A\to\infty]{} \frac{1}{1-t} \quad \text{for } t < 1. \tag{3.11}$$

Obviously,

$$M_A^{H_A}(t) = \mathbf{E}\mathbf{E}\left(e^{t p_A N_A^{H_A}} | X(0)\right),$$

where $X(0)$ has distribution H_A. It follows that for every initial state $x \geq 0$ and all $t < 1$ the value of $p_A N_A^x$ has a moment generating function

$$M_A^x(t) = \mathbf{E} e^{t p_A N_A^x}$$

and

$$M_A^{H_A}(t) = \mathbf{E} M_A^{X(0)}(t) = \int_0^A M_A^x(t) H_A(dx).$$

For $t \leq 0$, by virtue of Theorem 1(i)

$$M_A^x(t) \xrightarrow[A\to\infty]{} \frac{1}{1-t}.$$

Let $0 < \varepsilon < 1$ and $C > 0$ be such that $H\{[0, C)\} = \varepsilon$. For fixed $0 < t < 1$, let $A(\varepsilon) > C$ be such that

$$1 - \varepsilon < \frac{M_A^{H_A}(t)}{1/(1-t)} < 1 + \varepsilon \quad \text{whenever } A \geq A(\varepsilon).$$

Recall that $X(0)$ has distribution H_A, which is a quasi-stationary distribution.

For any $0 < \gamma < \infty$, Markov's inequality yields

$$\mathbf{P}\left(M_A^{X(0)}(t) > \gamma M_A^{H_A}(t)\right) \le 1/\gamma,$$

so that for $A \ge A(\varepsilon)$

$$\mathbf{P}\left(M_A^{X(0)}(t) > \frac{\gamma}{1-t}\right) \le \frac{1+\varepsilon}{\gamma}. \tag{3.12}$$

Substituting $\gamma = (1+\varepsilon)/\varepsilon$ in (3.12) yields

$$\mathbf{P}\left(M_A^{X(0)}(t) > \frac{1+\varepsilon}{\varepsilon}\frac{1}{1-t}\right) \le \varepsilon.$$

Since, by Lemma 6, $\varepsilon = H\{[0,C)\} \le H_A\{[0,C)\}$, it follows that for $M_A^{X(0)}(t) \ge \frac{1+\varepsilon}{\varepsilon}\frac{1}{1-t}$, the value of $X(0)$ cannot exceed C. In other words,

$$M_A^x(t) \le \frac{1+\varepsilon}{\varepsilon}\frac{1}{1-t} \quad \text{for } x \ge C \text{ and all } A \ge A(\varepsilon). \tag{3.13}$$

Let $\beta = \min\{n : X(n) \ge C\}$. Obviously,

$$M_A^0(t) = \mathbf{E}e^{tp_A N_A^0} \le \mathbf{E}e^{tp_A\beta} \cdot \mathbf{E}e^{tp_A N_A^C}. \tag{3.14}$$

Let $\delta_\varepsilon = \mathbf{P}\{X^0(1) \ge C\}$. Clearly $\delta_\varepsilon \to \mathbf{P}\{X^0(1) > 0\} > 0$ as $\varepsilon \to 0$.

Due to the monotonicity of the process $X(n)$, β is bounded by a $\mathrm{Geometric}(\delta_\varepsilon)$-distributed random variable, so that for $0 < t < 1$

$$1 \le \mathbf{E}e^{tp_A\beta} \le \mathbf{E}e^{tp_A \mathrm{Geometric}(\delta_\varepsilon)} = \frac{\delta_\varepsilon e^{p_A t}}{1-(1-\delta_\varepsilon)e^{p_A t}}.$$

It follows that $\mathbf{E}e^{tp_A\beta}$ is bounded as $A \to \infty$ (since $p_A \xrightarrow[A\to\infty]{} 0$). Since $\mathbf{E}e^{tp_A N_A^C} = M_A^C(t)$, equations (3.13) and (3.14) imply that $M_A^0(t)$ is also bounded as $A \to \infty$.

Denote $\varphi(t) = \limsup_{A\to\infty} M_A^0(t) < \infty$. Let $\{A_i\}_{i=1}^\infty$ be a sequence such that $\lim_{i\to\infty} M_{A_i}^0(t) = \varphi(t)$. Construct a set $\{t_j\}_{j=1}^\infty$ dense in $(0,t)$. Because $M_A^0(u)$ is monotone in u, one can obtain a subsequence $\{A_{ij}\}$ of $\{A_i\}$ such that $M_{A_{ij}}^0(u)$ converges as $j \to \infty$ for all $0 < u < t$. Since the limit is a moment generating function, by Theorem 1(i) it must be $1/(1-t)$. The same argument can be applied to $\liminf_{A\to\infty} M_A^0(t)$.

It follows that the limit $\lim_{A\to\infty} M_A^0(t)$ exists and is equal to $1/(1-t)$ for all $t < 1$. Because $M_A^x(t)$ is monotone in x and because of (3.11), $\lim_{A\to\infty} M_A^x(t)$ necessarily equals $1/(t-1)$ for all $t < 1$ and every fixed $x \in [0,\infty)$. This completes the proof of Theorem 1(ii).

4. Remarks

1. Let G be a distribution with support $[0,A]$ and define the operator T as

$$T(G) = \text{the distribution of } X(1) \text{ conditioned on } \{X(1) \le A, X(0) \sim G\}.$$

If T is a continuous operator (in the weak* topology on the distribution functions over $[0, A]$), then a quasi-stationary distribution exists, i.e., Condition C2 in Theorem 1 is satisfied (cf. Harris, 1963, Theorem III.10.1).

2. Even if T is not a continuous operator, sometimes Condition C2 can be verified by solving for $T(G) = G$ and arguing that this is the quasi-stationary distribution. For an example, see Pollak (1985).

3. The proof can be modified easily to extend Theorem 1 to the case where the support of the stationary distribution H is $[c, \infty)$ for some $c > 0$ (i.e., the set $[0, c)$ is not in the state space or is transient).

5. Examples and Applications

Theorem 1 can be applied to a number of popular Harris recurrent Markov processes. Below we present two examples. These are of interest when applying certain change-detection procedures.

5.1. Example 1: An Additive-Multiplicative Markov Process

Let $\Lambda_1, \Lambda_2, \dots$ be non-negative continuous independent and identically distributed (i.i.d.) random variables with $\beta = \mathbf{E}\Lambda_i$ and $\mu = \mathbf{E}\log\Lambda_i$. For $x \geq 0$, define recursively:

$$X(0) = x, \quad X(n) = (1 + X(n-1))\,\Lambda_n, \quad n = 1, 2, \dots\, . \tag{5.1}$$

This process is of interest in a number of applications (cf. Kesten, 1973; Pollak, 1985, 1987). For example, in the problem of detecting a change in distribution, the Shiryaev-Roberts statistic can be written as (cf. Pollak, 1985, 1987)

$$R(n) = (1 + R(n-1))\frac{f_{\theta_1}(Y_n)}{f_{\theta_0}(Y_n)}, \quad R(0) = 0, \tag{5.2}$$

where $\{Y_n, n \geq 1\}$ are independent, having probability density f_{θ_0} before a change and putative density f_{θ_1} after a change; θ_0 and θ_1 are fixed parameters, and one stops and declares that the change is in effect at $N_A = \min\{n : R(n) > A\}$.

When $\mu < 0$, the process $\{X(n)\}$ is Harris-recurrent and has a stationary distribution (for any $x \geq 0$). To see this, note that $X(n)$ can be written as

$$X(n) = \sum_{k=0}^{n}\prod_{i=k}^{n}\Lambda_i = \sum_{k=0}^{n}\exp\left\{\sum_{i=k}^{n}\log\Lambda_i\right\},$$

where $\Lambda_0 = x$. Obviously,

$$\sum_{k=0}^{n}\exp\left\{\sum_{i=k}^{n}\log\Lambda_i\right\} \overset{dist}{=} \sum_{k=1}^{n}\exp\left\{\sum_{i=1}^{k}\log\Lambda_i\right\} + x\exp\left\{\sum_{i=1}^{n}\log\Lambda_i\right\},$$

where the right hand-side converges (for every $x \geq 0$ as $n \to \infty$) to the random variable

$$\sum_{k=1}^{\infty}\exp\left\{\sum_{i=1}^{k}\log\Lambda_i\right\},$$

which is a.s. finite when $\mu < 0$. Since we assumed above that Λ_1 is continuous, the quasi-stationary distribution exists (see Remark 1 in Section 4). It follows from Theorem 1 that a suitably standardized version of the first exceedance time over A (i.e., $p_A N_A^x$) is asymptotically exponentially distributed.

Note that while using the conventional regeneration argument is perhaps possible, embedding the Markov process (5.1) into "regenerative cycles" by no means is either straightforward or obvious, which is especially true when $1 \le \beta = \mathbf{E}\Lambda_i < \infty$ and $\mu = \mathbf{E}\log\Lambda_i < 0$. This case does have meaning for applications. For example, regard the aforementioned change detection problem. When there never is a change, the observations $Y_i, i \ge 1$ have density f_{θ_0}, so that $\beta = \int [f_{\theta_1}(y)/f_{\theta_0}(y)] f_{\theta_0}(y)dy = 1$ while by Jensen's inequality $\mu = \int \log[f_{\theta_1}(y)/f_{\theta_0}(y)] f_{\theta_0}(y)dy < 0$. If there is a change – for argument's sake let it be in effect from the very beginning – the observations $Y_i, i \ge 1$ have density f_θ (not necessarily f_{θ_1}; the post-change parameter is seldom known in advance, and the putative θ_1 is merely a representation of a "meaningful" change). For θ close to θ_0, one would obtain $\beta = \int [f_{\theta_1}(y)/f_{\theta_0}(y)] f_\theta(y)dy > 1$ and $\mu = \int \log[f_{\theta_1}(y)/f_{\theta_0}(y)] f_\theta(y)dy < 0$.

Before going into further details, we discuss an issue related to computing p_A, the standardizing factor. If p_A were amenable to direct calculation, one could use this to approximate $\mathbf{E}N_A^x \approx 1/p_A$. Unfortunately, in most cases direct evaluation of p_A is not tractable, and evaluation of $\mathbf{E}N_A^x$ has to be done by other methods. (But see Pollak, 1985, and Mevorach and Pollak, 1991 for examples that allow some tractability.) Nonetheless, evaluation of p_A is of interest on its own merits (cf. Tartakovsky, 2005), as p_A is an approximation of the probability that there will be a first upcrossing of the threshold A at a specified time n, and $1-(1-p_A)^m$ is an approximation of the probability that there will be a first upcrossing of A in a given stretch of m observations (i.e., for the "local false alarm probability" $\mathbf{P}(n \le N_A^x \le n+m-1 | N_A^x \ge n)$). Therefore, if $\mathbf{E}N_A^x$ can be evaluated, p_A can be approximated by $1/\mathbf{E}N_A^x$.

Suppose now that $\beta = \mathbf{E}\Lambda_i = 1$. Let f_0 be the density of Λ_i and define $f_1(\Lambda) = \Lambda f_0(\Lambda)$. (Since $\mathbf{E}\Lambda = 1$, it follows that f_1 is a bona fide probability density.) Note that Λ is a likelihood ratio, $\Lambda = f_1(\Lambda)/f_0(\Lambda)$. It follows from Pollak (1987) (see also Tartakovsky and Veeravalli, 2005) that

$$\mathbf{E}_{f_0} N_A^x = \gamma^{-1} A(1+o(1)) \quad \text{as } A \to \infty, \tag{5.3}$$

where $\mathbf{E}_{f_0}$ is the expectation with respect to the density f_0 and γ is a constant that can be calculated by renewal theory (cf. Woodroofe, 1982; Siegmund, 1985), so that $p_A \approx \gamma/A$. See Remark in the end of Section 5.2 for evaluation of p_A when $\mathbf{E}\Lambda_i \ne 1$.

5.2. Example 2: A Reflected Random Walk

Let $\{Z_n\}_{n=1}^\infty$ be a sequence of i.i.d. continuous random variables with a negative mean $\mu = \mathbf{E}Z_n < 0$. For $n \ge 1$, define

$$X(n) = \max\{0, X(n-1) + Z_n\}, \quad X(0) = x \ge 0. \tag{5.4}$$

Since $\mu < 0$, the Markov process $\{X(n)\}$ is Harris-recurrent and has a stationary distribution. To see this, note that

$$X(n) = \max\{0, Z_1 + \cdots + Z_n + x, Z_2 + \cdots + Z_{n-1}, \ldots, Z_n\}.$$

Write $S_i = \sum_{k=1}^i Z_k$, $S_0 = 0$. Since the vector $(Z_1, \ldots, Z_n)$ has the same distribution as $(Z_n, \ldots, Z_1)$, it follows that

$$X(n) \overset{dist}{=} \max\{\max\{0, S_1, S_2, \ldots, S_{n-1}\}, x + S_n\},$$

where the right hand-side converges (as $n \to \infty$ for any $x \geq 0$) to the random variable $\max_{i\geq 0} S_i$, which is a.s. finite whenever $\mu = \mathbf{E}Z_i < 0$.

The process (5.4) describes a broad class of single-channel queuing systems (see, e.g., Borovkov, 1976) as well as a popular cumulative sum decision statistic for detecting a change in distribution (Page, 1954) and has been studied extensively by itself, outside the framework of general Markov processes. For instance, for $x = 0$, the asymptotic exponentiality of the stopping time

$$N_a^{x=0} = \min\{n \geq 1 : X(n) > a\}, \quad a > 0 \tag{5.5}$$

(as $a \to \infty$) has been proven by Khan (1995), which can be generalized easily for any $x > 0$. (The process $\{X(n)\}$ obviously is a renewal process, so, although our Theorem 1 covers this example when the conditions C1 and C2 are satisfied, it is not needed to prove asymptotic exponentiality of N_A^x, as it can be derived from general results; cf. Asmussen, 2003, Ch. VI.)

Assume for simplicity that $x = 0$. If there exists a positive ω such that $\mathbf{E}e^{\omega Z_i} = 1$, let $f_0(z)$ be the density of Z_i and define $f_1(z) = e^{\omega z} f_0(z)$. Since $\mathbf{E}e^{\omega Z_i} = 1$, it follows that f_1 is a bona fide probability density, and $f_1(Z)/f_0(Z) = e^{\omega Z}$ is a likelihood ratio. Hence, assuming that $\mu_1 = \int \log[f_1(z)/f_0(z)] f_1(z)dz < \infty$ and letting

$$N_a^0 = \min\{n \geq 1 : \max(0, \omega X(n-1) + \omega Z_n) > \omega a\},$$

standard renewal-theoretic methods (cf. Woodroofe, 1982; Siegmund, 1985) readily apply to obtain that

$$\mathbf{E}N_a^0 = \delta^{-1} e^{\omega a}(1 + o(1)) \quad \text{as } a \to \infty, \tag{5.6}$$

so that $p_A \approx \delta e^{-\omega a}$. Here $0 < \delta < 1$ is a constant that can be computed explicitly by a renewal-theoretic argument (cf. Tartakovsky, 2005).

Remark. Clearly, N_a^x of Example 2 is larger than N_A^x of Example 1 (with $A = e^a$), so that $\mathbf{E}N_A^x \leq \delta^{-1} A^{\omega}(1 + o(1))$. Theorem 5 of Kesten (1973) as well as Theorem 4 of Borovkov and Korshunov (1996) imply that

$$\mathbf{P}(X(\infty) > y) = C/y^{\omega}(1 + o(1)) \quad \text{as } y \to \infty,$$

where $X(\infty)$ is a random variable that has the stationary distribution of $\{X(n)\}$ and C is a positive finite constant. Note that $X(\infty)$ is stochastically larger than a random variable that has the quasi-stationary distribution. Therefore, the first upcrossing over A of the process $X(n)$ starting at a random $X(0)$ distributed like $X(\infty)$ will occur no later than the first upcrossing over A of the process $X(n)$ starting at a random $X(0)$ that has the quasi-stationary distribution. The proportion of times that the former exceeds A is $\mathbf{P}(X(\infty) > A)$. It follows that $\mathbf{E}N_A^x \geq C^{-1} A^{\omega}(1 + o(1))$, so that p_A has an order of magnitude $1/A^{\omega}$.

5.3. Applications to Sequential Change-Point Detection and a Monte Carlo Experiment

The importance of the asymptotic exponentiality of the run length in sequential change-point detection methods is twofold. First, it shows that the mean time to false alarm (the so-called average run length), which is a popular measure of the false alarm rate, is indeed an exhaustive performance metric. Second, the result can be used for the evaluation of the local false alarm probabilities of the corresponding detection schemes (see Example 1 above; see Tartakovsky (2005)

for a more detailed discussion of the importance of local false alarm probabilities in a variety of applications).

To be more specific, assume that there is a sequence i.i.d. variables (observations) $Y_1, Y_2, \ldots$ that follow the density $f_0(y)$ under the no-change hypothesis (the in-control mode) and the density $f_1(y)$ after the change occurs (the out-of-control mode). The change occurs at an unknown point in time ν; $1 \le \nu < \infty$. Therefore, conditioned on $\nu = k$, $Y_n \sim f_0(y)$ for $n < k$ and $Y_n \sim f_1(y)$ for $n \ge k$. We write $\mathbf{P}_\infty$ ($\mathbf{E}_\infty$) and $\mathbf{P}_k$ ($\mathbf{E}_k$) respectively for probability measures (expectations) when there is no change (i.e., $\nu = \infty$) and when the change occurs at point k. Let $Z_n = \log[f_1(Y_n)/f_0(Y_n)]$ be the corresponding log-likelihood ratio and let $S_n = \sum_{i=1}^n Z_i$. Let $I_1 = \mathbf{E}_1 Z_1$ and $I_0 = \mathbf{E}_\infty(-Z_1)$ be the Kullback-Leibler information numbers, which are assumed finite.

We begin with the cumulative sum (CUSUM) test. The CUSUM statistic is given by the recursion (5.4) and the corresponding stopping time is defined in (5.5). The difference from the previous section is that Z_n, $n = 1, 2, \ldots$ are not arbitrary random variables with negative mean, but rather log-likelihood ratios with mean $\mu = -I_0$. This simplifies most of the calculations, since $\mathbf{E} e^{Z_n} = 1$. Recall that in this section we denote this expectation by $\mathbf{E}_\infty$.

Rewrite the corresponding stopping time in the following form

$$N_A = \min\left\{n \ge 1 : \max\left\{1, W(n-1) + e^{Z_n}\right\} > A\right\}, \tag{5.7}$$

where $W(0) = 1$ and $A = e^a$. The asymptotic approximation for the average run length to false alarm (5.6) holds with $\omega = 1$, $e^a = A$, and $\delta = I_1\gamma^2$ (cf. Tartakovsky, 2005), which implies that $p_A \sim I_1\gamma^2/A$. Here $\gamma = \lim_{y\to\infty} \mathbf{E}_1 \exp\{-(S_{\tau_y} - y)\}$, where $\tau_y = \min\{n : S_n > y\}$ is the first time when the random walk $S_n = \sum_{i=1}^n Z_i$ crosses the level y. The constant γ is the subject of renewal theory (cf. Woodroofe, 1982 or Siegmund, 1985) and can be computed explicitly.

We now proceed with the Shiryaev-Roberts detection test. The Shiryaev-Roberts statistic is defined by (5.2), where $\frac{f_{\theta_1}(Y_n)}{f_{\theta_0}(Y_n)} = e^{Z_n}$ and $R(0) = 0$. The corresponding stopping time is

$$\hat{N}_A = \min\left\{n \ge 1 : R(n) > A\right\}.$$

We now denote it by $\hat{N}_A$ to distinguish from the CUSUM stopping time in the following calculations and comparison.

Since $\mathbf{E}_\infty e^{Z_n} = 1$, the process $R(n) - n$ is a zero-mean martingale, which allows us to approximate the average run length to false alarm:

$$\mathbf{E}_\infty \hat{N}_A \sim \gamma^{-1} A \quad \text{as } A \to \infty.$$

This approximation follows from (5.3) above. The distribution of the Shiryaev-Roberts stopping time is approximately $\mathrm{Exponential}(p_A)$ with $p_A \sim \gamma/A$. (The asymptotic exponentiality of the suitably standardized run length to false alarm has been shown by Yakir, 1995.)

In order to verify the accuracy of asymptotic approximations for reasonable values of the threshold A, we performed Monte Carlo (MC) simulations for the following example. Consider the case where observations are independent, originally having an $\mathrm{Exponential}(1)$ distribution, changing at an unknown time to $\mathrm{Exponential}(1/(1+q))$, i.e.,

$$f_0(y) = e^{-y}\mathbb{1}_{\{y\ge 0\}}, \quad f_1(y) = \frac{1}{1+q} e^{-y/(1+q)}\mathbb{1}_{\{y\ge 0\}}, \quad q > 0. \tag{5.8}$$

In this case

$$I_1 = q - \log(1+q) \quad \text{and} \quad \gamma = 1/(1+q).$$

Applying Example 1, the likelihood ratio is $\Lambda_n = e^{Z_n} = (1+q)^{-1} e^{qY_n/(1+q)}$ and the average run length (ARL) to false alarm of the Shiryaev-Roberts procedure is

$$\mathrm{ARL}_{SR}(A) = \mathbf{E}_\infty \hat{N}_A \approx (1+q)A. \tag{5.9}$$

Applying Example 2, an approximation of the ARL to false alarm of the CUSUM test is

$$\mathrm{ARL}_{CU}(A) = \mathbf{E}_\infty N_A \approx \frac{(1+q)^2}{q - \log(1+q)} A. \tag{5.10}$$

Table 1: The ARL versus threshold for the CUSUM test for $q = 3$

A	1.2	1.7	2.5	4.6	9.2	13.0	17.1	21	41
FO ARL_{CU}	11.90	16.86	24.79	45.61	91.22	128.90	169.55	208.22	406.52
HO ARL_{CU}	7.96	12.36	19.69	39.56	84.07	121.21	161.43	199.77	397.02
MC $\widehat{\mathrm{ARL}}_{CU}$	8.04	12.45	19.79	39.57	84.33	121.23	161.88	200.44	397.16
MC $\mathrm{SD}(N_A)$	7.49	11.88	19.18	38.61	83.21	119.73	159.91	198.97	396.84

Table 2: The ARL versus threshold for the Shiryaev-Roberts test for $q = 3$

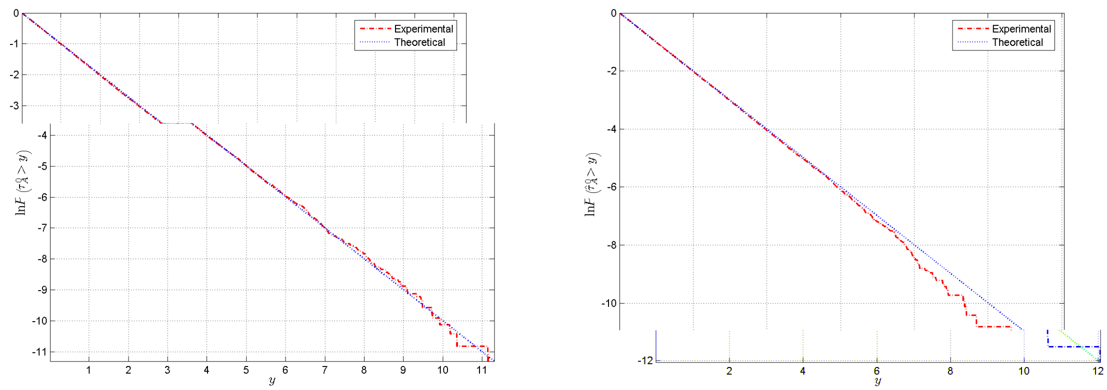
A	1	2	5	10	20	30	40	50	100
ARL_{SR}	4	8	20	40	80	120	160	200	400
MC $\widehat{\mathrm{ARL}}_{SR}$	4.01	8.03	20.00	39.94	79.99	119.82	159.17	200.42	399.46
MC $\mathrm{SD}(\hat{N}_A)$	3.00	6.78	18.34	37.92	77.33	117.39	157.19	197.90	396.94

We simulated the CUSUM and Shiryaev-Roberts procedures under the assumption of no change (i.e., all simulated observations are $\mathrm{Exponential}(1)$). Each combination of (test,threshold) was simulated 100,000 times. The results are reported in Tables 1 and 2. We present the results of simulations when the parameter $q = 3$, which is a reasonable value in certain applications such as detection of a randomly appearing target in noisy measurements, in which case q is the signal-to-noise ratio (see, e.g., Tartakovsky, 1991 and Tartakovsky and Ivanova, 1992). It is seen that the approximation (5.9) for the Shiryaev-Roberts test is very accurate for all threshold values, even when the ARL is small. On the other hand, the approximation (5.10) for the CUSUM test (given in the row "FO ARL_{CU}" in Table 1, where FO stands for "first order") is not especially accurate. This happens primarily because the first order approximation takes into account only the first term of expansion and ignores the second term $O(\log A)$ as well as constants. An accurate, higher order (HO) approximation can be obtained using the results of Tartakovsky and Ivanova (1992) which give:

$$\begin{aligned}\mathrm{ARL}_{CU}(A) \approx \frac{(1+q)^2}{q - \log(1+q)} A - \frac{1}{\log(1+q) - q/(1+q)} \log A \\ - \frac{1+q}{q - \log(1+q)} - \frac{q}{(1+q)\log(1+q) - q}.\end{aligned}$$

In Table 1, the row "HO ARL_{CU}" corresponds to this latter approximation, which perfectly fits the MC estimates (denoted by "MC $\widehat{\mathrm{ARL}}_{CU}$") for all tested threshold values $A \geq 1.2$.

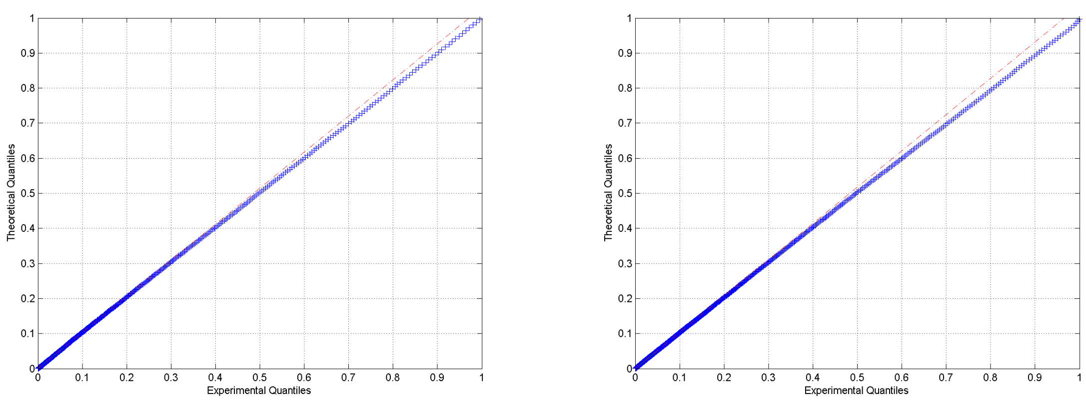
In these tables we also present the MC estimates of standard deviations $\mathrm{SD}(N_A)$ and $\mathrm{SD}(\hat{N}_A)$ of the stopping times. As one would expect, the standard deviations are the same (approximately) as the means, and the similarity grows as A increases. The fit is slightly better for the CUSUM test.



(a) CUSUM test: $q = 3$, $A = 13$

(b) Shiryaev-Roberts test: $q = 3$, $A = 40$

Figure 1: Empirical estimates of $\log[\mathbf{P}_\infty(\tau_A > y)]$ and $\log[\mathbf{P}_\infty(\hat{\tau}_A > y)]$ for the CUSUM and Shiryaev-Roberts procedures



(a) CUSUM test: $q = 3$, $A = 13$

(b) Shiryaev-Roberts test: $q = 3$, $A = 40$

Figure 2: QQ-plots for the stopping times of the CUSUM and Shiryaev-Roberts procedures

Figures 1(a) and 1(b) show the logarithm of the empirical (MC estimates) survival functions $\log \mathbf{P}_\infty(\tau_A > y)$ and $\log \mathbf{P}_\infty(\hat{\tau}_A > y)$ for the CUSUM and Shiryaev-Roberts procedures, where $\tau_A = N_A/\widehat{\mathrm{ARL}}_{\mathrm{CU}}$ and $\hat{\tau}_A = \hat{N}_A/\widehat{\mathrm{ARL}}_{\mathrm{SR}}$ are the corresponding standardized stopping times, along with the logarithm of the exponential probability plot $\log e^{-y} = -y$. The quantile-quantile plots (QQ-plots) for the stopping times are shown in Figures 2(a) and 2(b). The QQ-plots display

sample quantiles of N_A and $\hat{N}_A$ versus theoretical quantiles from the exponential distribution. If the distributions of the stopping times are exponential, the plots will be close to linear. These figures show that, for the chosen putative value of the post-change parameter ($q = 3$), the exponential distribution approximates the distributions of the stopping times very well. It is seen that the exponential approximation works very well already for $A = 13$ ($\mathrm{ARL}_{\mathrm{CU}} \approx 120$) for the CUSUM test and for $A = 40$ ($\mathrm{ARL}_{\mathrm{SR}} \approx 160$) for the Shiryaev-Roberts test. When considering that in practical applications the values of the ARL to false alarm usually range from 300 and upwards, the exponential distribution seems to be a perfect fit.

Acknowledgements

We are grateful to Alexei Polunchenko for the help with Monte Carlo simulations.

Moshe Pollak is Marcy Bogen Professor of Statistics at the Hebrew University of Jerusalem. His work was supported in part by a grant from the Israel Science Foundation, by the Marcy Bogen Chair of Statistics at the Hebrew University of Jerusalem, and by the U.S. Army Research Office MURI grant W911NF-06-1-0094 at the University of Southern California. The work of Alexander Tartakovsky was supported in part by the Marcy Bogen Chair of Statistics at the Hebrew University of Jerusalem and by the U.S. Office of Naval Research grant N00014-06-1-0110 and the U.S. Army Research Office MURI grant W911NF-06-1-0094 at the University of Southern California.

References

[1] ASMUSSEN, S. (2003). *Applied Probability and Queues*, Second Edition. SpringerVerlag, New York.

[2] BOROVKOV, A.A. (1976). *Stochastic Processes in Queuing Theory*. Springer-Verlag, New York.

[3] BOROVKOV, A.A. AND KORSHUNOV, D.A. (1997). Large-deviation probabilities for one-dimensional Markov chains, Part 1: Stationary distributions. *Theory Probab. Appl.* **41**, No. 1, 1–24.

[4] GLASSERMAN, P. AND KUO S-G. (1995). Limits of first passage times to rare sets in regenerative processes. *Ann. Appl. Probab.* **5** 424–445.

[5] HARRIS, T.E. (1963). *The Theory of Branching Processes*. Springer Verlag, Berlin.

[6] KESTEN, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131** 207–248.

[7] KHAN, R.A. (1995). Detecting changes in probabilities of a multi-component process. *Sequential Analysis.* **14** 375–388.

[8] MEVORACH, Y. AND POLLAK, M. (1991). A small sample size comparison of the Cusum and Shiryayev-Roberts approach to changepoint detection. *The American Journal of Mathematical and Management Sciences.* **11** 277–298.

[9] PAGE, E.S. (1954). Continuous inspection schemes. *Biometrika.* **41** 100–115.

[10] POLLAK, M. (1985). Optimal detection of a change in distribution. *Ann. Statist.* **13** 206–227.

[11] POLLAK, M. (1987). Average run lengths of an optimal method of detecting a change in distribution. *Ann. Statist.* **15** 749–779.

[12] POLLAK, M. AND SIEGMUND, D. (1986). Convergence of quasi-stationary to stationary distributions for stochasically monotone Markov processes. *J. Appl. Prob.* **23** 215–220.

[13] SIEGMUND, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Springer-Verlag, New York.

[14] TARTAKOVSKY, A.G. (1991). *Sequential Methods in the Theory of Information Systems*. Radio and Communications, Moscow (In Russian).

[15] TARTAKOVSKY, A.G. (2005) Asymptotic performance of a multichart CUSUM test under false alarm probability constraint. *Proc. 44th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC'05)*, December 12-15, pp. 320–325, Seville, Spain, Omnipress CD-ROM.

[16] TARTAKOVSKY, A.G. AND IVANOVA, I.A. (1992). Comparison of some sequential rules for detecting changes in distributions. *Problems Inform. Transm.* **28** 117–124.

[17] TARTAKOVSKY, A.G. AND VEERAVALLI, V.V. (2005). General asymptotic Bayesian theory of quickest change detection. *Theory Probab. Appl.* **49** 458-497.

[18] WOODROOFE, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*. SIAM, Philadelphia.

[19] YAKIR, B. (1995). A note on the average run length to false alarm of a change-point detection policy. *Ann. Statist.* **23** 272–281.